SCIEX
The Power of Precision

Fast and efficient identification of candidate biotherapeutics through high-throughput screening analysis

Supporting biopharma development with QTOF technology and Biologics Explorer software

Ebru Selen, Zoe Zhang and Kerstin Pohl
Biologics
SCIEX, USA

In this technical note, a workflow for high-throughput (HT) screening of intact and subunit masses for identification purposes of candidate biotherapeutic proteins using Biologics Explorer software is showcased.

Liquid chromatography coupled to MS (LC-MS) instrumentation is at the forefront of biopharmaceutical development processes, enabling researchers to identify, characterize and investigate dozens of different candidates to meet the growing demand for biotherapeutics. Although QTOF MS offers excellent sensitivity and mass accuracy, perfectly suited for the identification of intact and subunit proteins, data analysis is often a bottleneck. Labor-intense, non-automated data analysis, such as, file-by-file deconvolution and matching to theoretical sequence information, pose challenges for MS utilization in early development protein screening. As a result, MS adoption is more focused on late early-development stages, after candidates had been narrowed down by other methodologies, providing less qualitative information. Being able to close the data processing gap can improve informed decision making and reduce the risk of not pursuing promising biotherapeutic candidate proteins.

Here, the challenges of LC-MS data processing of early development protein screenings are addressed using Biologics Explorer software. Intact and subunits of several monoclonal antibodies (mAb) were used to investigate the suitability for molecular weight assessment of different potential candidates. The import of a vast number of data files, automatic deconvolution and intuitive reporting with minimal user input offer a solution to existing processing challenges. In addition, transparency is provided by access to all data processing steps as needed.

Key features of the protein screening workflow

- Increased efficiency in screening of biotherapeutic proteins on intact and subunit level through a simple HT workflow set-up
- Minimal user input to screen hundreds of biotherapeutic candidates in a single, efficient processing step
- Easy-to-read, visually supported report outputs facilitate decision making in fast-paced biopharma environments
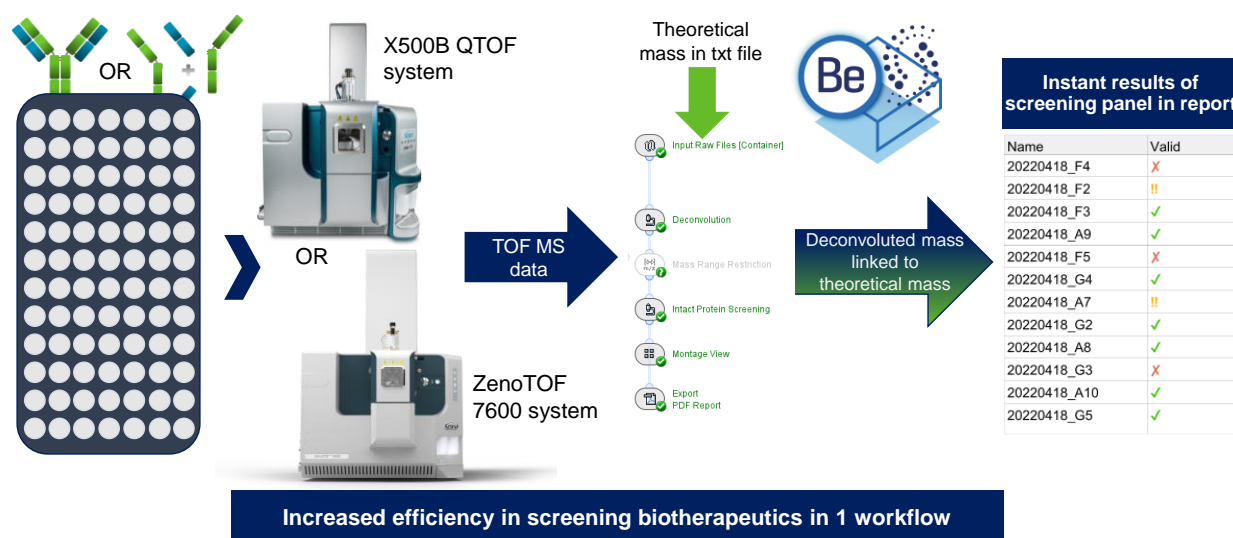

Figure 1. HT biotherapeutic screening workflow. Screening of the MW of hundreds of different biotherapeutic proteins in a fast and efficient manner can be achieved with a user-friendly workflow in Biologics Explorer software. The report provides an intuitive overview of which samples matched the theoretical MW information (green check mark), those that had showed marginal differences in terms of mass accuracy (yellow exclamation points) and which samples above the set threshold for MW matching (red x).

Methods

Sample preparation: Candidate biotherapeutics were diluted in 50mM Tris to a working concentration of 0.25 µg/µL. For subunit analysis, 150 µL denaturing reagent (7.2M Guanidine in 50mM Tris, pH 7-8) was added to 50 µL sample and vortexed. DTT was added to a final concentration of 100mM and mixture was incubated at 60ºC for 30 mins. Samples were diluted with 0.1% formic acid by a factor of 2 and transferred into vials. For intact analysis, mAbs in working concertation were transferred to the vials.

Chromatography: For reversed-phase liquid chromatography (RPLC) of intact and subunit samples, a Waters BioResolveRP, mAb Polyphenyl, column (2.1 mm x 50 mmx 2.7 µm) was used. Flow rate was set to 0.25 mL/min, 2 µL of sample was injected and the mobile phase A consisted of 0.1% formic acid in water and mobile phase B of 0.1% formic acid in acetonitrile. Table 1 shows gradient for intact and subunit analysis.

Table 1. Chromatography for intact and subunit analysis.

Time [min]	Mobile phase A [%]	Mobile phase B [%]
Initial	73	27
2	73	27
15	55	45
16	10	90
18	10	90
18.1	90	27
20	90	27

Mass spectrometry: Data were acquired in the TOF MS positive ionization mode using a ZenoTOF 7600 system with intact protein mode enabled. Detailed MS parameters are shown in Table 2.

Table 2. MS parameters for subunit and intact mass analysis.

Parameter	Setting
Scan mode	TOF MS
Polarity	Positive
Intact protein mode	ON
Gas 1	50 psi
Gas 2	50 psi
Curtain gas	37 psi
Temperature	500℃
Ion spray voltage	5500 V
CAD gas	6
Time bins to sum	80
Accumulation time	0.5 s
Start mass	400 m/z
Stop mass	4000 m/z
Declustering potential	200 V
Collision energy	10 V

Data processing: Data were processed and visualized using Biologics Explorer software version 1.0.2. Biologics Explorer software is compatible with all SCIEX accurate mass spectrometers and data in .WIFF and .WIFF2 format.

From data files to results in an efficient manner

The mAb samples mimicking different biotherapeutic candidates were loaded into 96-well plate and TOF MS data were acquired using a generic LC-MS method setup. The LC and MS methods allow for the detection of different mAb species and their subunits with high data quality, skipping sample-specific optimization to address the need for fast and efficient data acquisition. To understand if the proteins were expressed as desired, raw TOF MS data needed to be deconvoluted and obtained molecular weights (MW) matched against theoretical sequence information. To initiate the processing workflow of MW determination and matching, all data files (each corresponding to a different protein sample) were imported to Biologics Explorer software at once (Figure 2A). For processing with the goal of molecular weight determination, broad criteria were defined for the retention time (RT) range used for automatic peak detection, the deconvolution and for the MW output (Figure 2B). Since different proteins will have different hydrophobicities depending on their sequences and modifications, RTs are likely to shift. Automatic peak detection within a broad RT range is therefore critical to ensure an accurate deconvolution. Similarly, the MW is likely to be different between different protein samples. Defining a wide mass range for the deconvolution output ensures the applicability for the workflow for many different samples. To automatically determine the MW of each candidate protein, MS spectra within the detected peaks were averaged and deconvoluted for all candidates at once maintaining consistency in deconvolution (Figure 2B, bottom). The matching of the resulting MW to theoretical sequences is enabled through a text file, linking the MW output of each data file to a specific theoretical mass (Figure 2C). In addition, 2 thresholds for the flagging of results can be set to distinguish between marginal and significant differences between expected and experimental masses.

Current screening methodologies for intact and subunit proteins are suffering from lack of automation of data analysis. Although deconvolution of multiple files simultaneously is possible with most software options, users still face the need of manual investigation of results to determine matches of expected sequences with deconvoluted masses. In a fast-paced environment where there is the desire to screen hundreds of candidates per day, routinization of processes becomes necessary, yet challenging. Here, this challenge is addressed by a generic method, which links the theoretical MW to empirical data in an efficient way. The automatic linkage is initiated by a tab-delimited text file including the sample file names and theoretical MW information (Figure 2C). The software automatically creates a link between the theoretical info and empirical data using the sample file name. The difference in theoretical mass and detected mass, expressed in ppm, is compared against the validity and attention thresholds, which facilitate the interrogation of results based on intuitive icons (Figures 1 and 3). Figure 3 represents examples of the sample reports in PDF format for intact and subunit analyses of candidate biotherapeutics. A candidate will be associated with 1 of 3 options: valid with a green check mark if the difference between expected and detected mass is lower than the absolute validity threshold, critical with yellow exclamation marks when the difference is lower than the attention threshold, but higher than the validity threshold, or invalid with a red cross if the difference is higher than the attention threshold, indicating the need for further investigation.
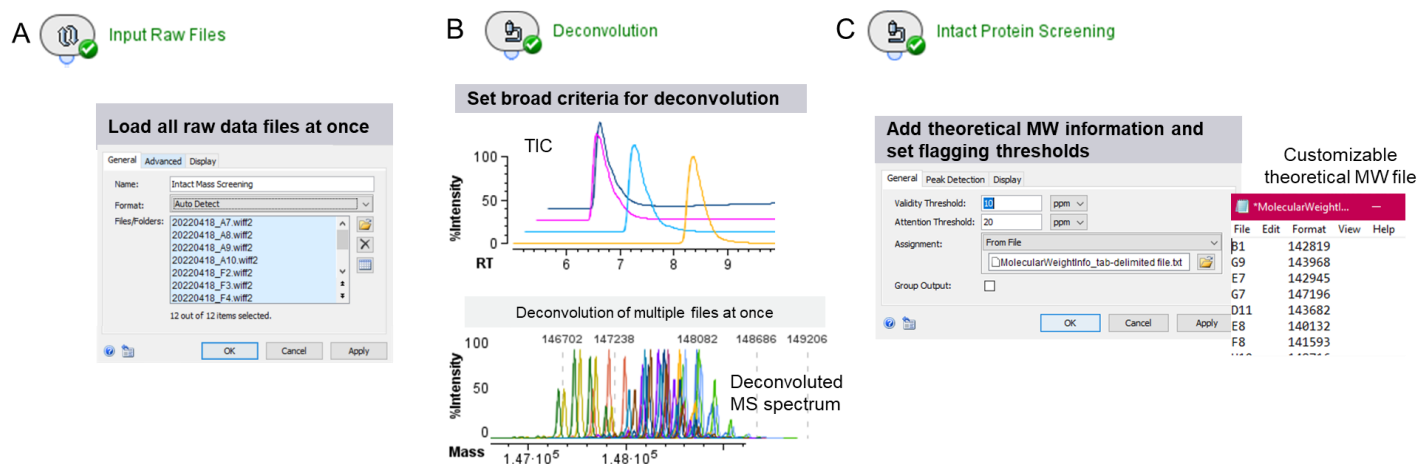

Figure 2. Screening a panel of biotherapeutics using Biologics Explorer software. Users upload data files (A). Peaks are automatically detected from total ion chromatograms (TIC) (B top panel), data are deconvoluted as a batch (B, bottom) using generic settings. Detected masses are automatically screened against a simple tab delimited text file including molecular weight information. In addition, flagging thresholds based on mass deviation of empirical vs. theoretical MW can be set for the report (C).

Export PDF Report — Ready-to-view screening results

Intact

Zoomed Range 1

Name	Expected Mass	Detected Mass	Delta [Da]	Delta [ppm]	Valid
20220418_F4	148082.17	148082.08	-0.09	-1.0	✓
20220418_F2	147237.13	147239.71	2.59	18.0	✓
20220418_F3	148348.17	148219.97	-128.2	-864.0	✗
20220418_A9	148080.9	148081.73	0.83	6.0	✓
20220418_F5	148205.0	148200.24	-4.76	-32.0	‼
20220418_G4	148080.9	148081.77	0.87	6.0	✓
20220418_A7	147237.13	147239.63	2.51	17.0	✓
20220418_G2	147239.0	147239.55	0.55	4.0	✓
20220418_A8	148219.9	148219.9	0.0	0.0	✓
20220418_G3	148200.0	148219.72	19.72	133.0	✗
20220418_A10	148200.0	148200.36	0.36	2.0	✓
20220418_G5	148200.0	148200.41	0.41	3.0	✓

Subunits

Zoomed Range 1

Name	Expected Mass	Detected Mass	Delta [Da]	Delta [ppm]	Valid
20220418_F7	23039.5	23039.58	0.08	3.0	✓
20220418_F8	23442.86	23443.08	0.22	9.0	✓
20220418_F9	23127.6	23127.67	0.07	3.0	✓
20220418_F10	23358.0	23412.02	54.02	2313.0	✗

Zoomed Range 2

Name	Expected Mass	Detected Mass	Delta [Da]	Delta [ppm]	Valid
20220418_F7	50677.3	50677.38	0.08	2.0	✓
20220418_F8	50602.4	50602.44	0.04	1.0	✓
20220418_F9	51069.7	51069.8	0.1	2.0	✓
20220418_F10	51928.8	50644.96	-1283.84	-24723.0	✗

Figure 3. Single button, ready-to-view results increase the efficiency in decision making in fast-paced environments. Results are provided in a customizable pdf format with information on sample name, expected and detected mass, mass errors and visual interpretations (last column) to efficiently interrogate results. Example shows results obtained from intact mass analysis on top, and light and heavy chain results from subunit analysis at the bottom.

The visual interpretation of the results with intuitive icons increases the efficiency in decision making in a fast-paced environment. Users can adjust the threshold settings to suit their needs (Figure 3). In addition to the file name, relevant information on expected and detected mass as well as the ppm deviation is shown. Furthermore, the information in the report is customizable.

Transparency in data analysis and results

The HT workflow is optimized to speed up the screening and decision-making process through user-friendly layouts, single-button executions and smart visual outputs. However, there might be the need to investigate a sample further, especially if a significant deviation from the expectation was found. The workflow therefore provides the option for data interrogation at each processing step for full transparency. All data is directly accessible in the software to facilitate any needed troubleshooting.

As an example, the data for the sample in well F3 were investigated (Figure 4). The red cross from the report of the rapid screening for sample F3 (Figure 3 and Figure 4A) is immediately capturing attention, reducing time spent on samples, which passed set criteria. A quick examination of the table in the report shows a ~128 Dalton difference between the expected mass and the detected mass of sample F3. Representative raw data and deconvoluted data were subsequently used to interrogate the data quality (Figure 4B): A peak with great S/N in the TIC indicates successful injection into the LC-MS system.
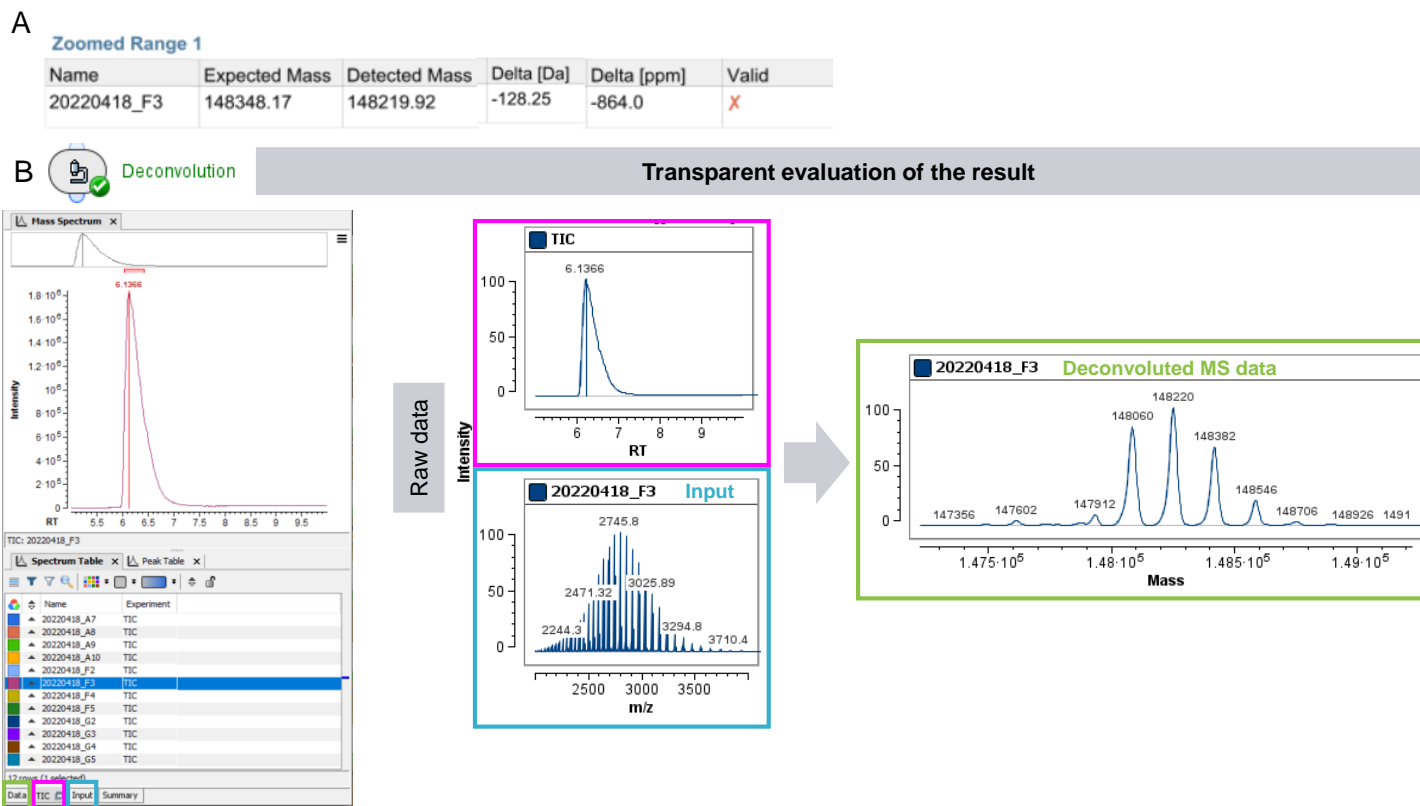
Figure 4. Demonstration of transparent data investigation in Biologics Explorer software. Sample F3 showed a mass deviation above the attention threshold in the report (A). Access to the TIC, TOF MS (input) and deconvoluted data (B) can be used to investigate the sample further.

Furthermore, a charge state envelope with baseline separation of proteoforms was achieved for the TOF MS data. The derived deconvoluted data shows as well a common pattern of a mAb. One explanation for the mismatch between expected and detected, deconvoluted mass could be a clipping event of an amino acid, such as lysine, leading to a lower detected MW than the expected one. At this step, a decision whether to continue investigating the sample or focusing on the samples which passed the criteria can be made. To investigate the theory of a clipping event, further in-depth analysis with a peptide mapping analysis can be performed. With that approach the sequence deviations and modification(s) of the candidate biotherapeutic can be investigated.

In summary, biopharmaceutical development can be taken one step further by enabling fast and efficient LC-MS-based screening of intact and subunit proteins with Biologics Explorer software. The user-friendly workflow closes current data analysis gaps, while allowing a transparent data analysis experience. Equipped with visual interpretation of results and customizable report options, the HT workflow supports fast-paced decision-making processes where high numbers of candidates need to be screened to meet the demand in analytics of biotherapeutics.

Conclusions

- HT screening of candidate biotherapeutics in 96-well plate settings was achieved at intact and subunit level with a simple and effective set-up in Biologics Explorer software leveraging accurate mass QTOF MS data

- Automatic batch processing and deconvolution of candidates within wide mass ranges and batch reporting allow routinization of QTOF MS-based screening of proteins during early development

- Smart visualization icons offered by the workflow provide easy-to-read reports and fast decision making for hundreds of candidate molecules